

Weekly Report

Lu Junhua

2015 年 7 月 19 日

This week, we went to Zhongshan Building 3 days/times. We had 3 reports for these visits. Here are the brief summary.

13th July(All 3):

- identify two ways of extract data.
- about the security of data.
- our concern about the complexity of social relationship under one 户号.
- an arrangement of the later work.

15th July(Lu Junhua):

- get familiar with the query code.
- write a function of searching people under same 户号.
- discovered bugs in displaying hotel records and revised it.
- permanent or temporary tag, occupations tag remains to be add.
- the way to store the data needs discussion.
- Prof.Gu's suggestions:

We wouldn't worry about categorized data, we have specific method for reducing the categories.

RARE EVENT: here are only very small amount of criminals. While using logistic regression, we have to adopt control-pairing method. 1:5 is a nice choice, we can also try 1:3 or 1:2.

We have to revise the estimation: $\text{logit}(Y)=a+bX$, for revising a: $a_{new} = a - \ln \frac{1-t}{t} \cdot \frac{\bar{y}}{1-\bar{y}}$. t is the real value of crime ratio. \bar{y} is the ratio of sample, as 1/5. b will not change.

15th July(Ke Junming):

- understanding the data structure.
- get the oracle database password. And ways to connect the database.

- find the 5 tables we need. as the following picture.

表名	表描述
HZ2004.HJXX_CZRKJBXXB	常住人口基本信息
NEWZZRK.JB_TAB	暂住人口基本信息
DFK_GAXZ.BADPERSON_S_CRIMINAL	犯罪 detail
LGY2007.LGT_RY_NB	离店时间，入住时间
WBXT.SWRY_XXB	网吧上网时间

- several missing and incorrects, do it next week.

Besides this, I learned basic syntax of JavaScript, jQuery, Html. Also learned a bit knowledge of the D3.js library. The implementation of several visualizations will be done next week. And also, I learned to build up my personal webpage using hexo and git.

And I went home on Friday.